



## Revolutionizing Cancer Research Through Informatics

### Moderator:

Welcome to the caBIG® podcast network. Part one of our podcast series describes how information technology is revolutionizing the field of cancer research. Joining us today is Dr. Ken Buetow, who's a geneticist by training. Dr. Buetow is Director of the National Cancer Institute's Center for Biomedical Informatics and Information Technology, and he leads the bioinformatics program that is linking cancer researchers across the country to share data and collaborate on research into this disease killer. Thanks for joining us today, Dr. Buetow.

### Dr. Buetow:

My pleasure.

### Moderator:

To get started, can you tell us what makes cancer such a challenge to understand?

### Dr. Buetow:

First, it's important to recognize that cancer is a disease of genes, and that it arises from the changes in the underlying molecular processes of our body. As scientists have begun in recent years to examine cancer at the molecular level, what we've recognized is that it's not a single disease, but in fact is made up of literally dozens, to perhaps hundreds of diseases. For instance, a cancer called acute lymphoblastic leukemia is made up of as many as 18 different diseases.

### Moderator:

And this complexity translates into the need to collect a great deal of data to understand the disease, correct?

### Dr. Buetow:

Yes. Because of this underlying complexity of cancer—the very complex etiology of so many different molecular underpinnings—we need to collect a tremendous amount of information. That information comes from profiling hundreds or even thousands of different individuals, as well as looking at the molecular characteristics of the disease itself.

That huge volume of data is well beyond the capacity of any single human being to hold in their head, and quite commonly, impossible even for an entire discipline—such as the cancer research community—to keep in their own filing cabinets or individual computers. So, what we need to do is use today's modern information technology tools to overcome those challenges in several ways.

First, we need to capture and manage a huge amount of complex information, and connect it across all of the different participants in the cancer community. Second, we need to be able to provide biomedical researchers with the capacity to connect the very different dimensions of molecular characterization of disease. Third, we need to put at the fingertips of practicing cancer physicians, the latest information about the treatments available, and how to distinguish the different sub-types of cancer I mentioned earlier; and, how to direct the therapy at those different types of cancer.



And, lastly, we need to empower patients and consumers not only to be in control of the information related to their health and disease, but also to have full access to the entire research endeavor, so they can participate in clinical trials and other forms of biomedical research.

**Moderator:**

And how is the National Cancer Institute doing all that information management?

**Dr. Buetow:**

The National Cancer Institute has launched a groundbreaking effort called caBIG®, or, the Cancer Biomedical Informatics Grid. caBIG® essentially applies 21st Century information technology to provide the cancer research and clinical care community with “electronic glue”—that is, the online connectivity necessary to bring together all of our efforts to combat cancer.

**Moderator:**

Dr. Buetow, can you describe caBIG® in greater detail?

**Dr. Buetow:**

My pleasure. caBIG® is fundamentally a network. It's a network of individuals, organizations, and institutions that have come together in a common effort to attack the troublesome disease known as cancer. But it's also a network in the technological sense in that it's a combination of people interconnecting, their basic computational infrastructure. It's also a network in the context of interconnected policies, across all of the different organizing, organizations and institutions working together as an integrated whole.

Let me be a little more specific about what I mean by the components of caBIG®. First, caBIG® is composed of information standards that allow different members of the community to share information in a standardized way, so that biologists and clinicians and pathologists and statisticians and all the other participants can connect and interpret the information that's part of the entire cancer enterprise.

Second, caBIG® is a collection of applications, of software tools that support the collection, storage, analysis and interpretation of this diverse and complex cancer information that we were just discussing a moment ago. There are, at this point in time, more than 40 caBIG® applications to support very diverse needs across the entire spectrum of cancer research.

Third (and perhaps of most importance) caBIG® is a basic interconnectivity infrastructure—an extension of the worldwide web and the fundamental Internet—that allows these different individuals and organizations and institutions to use the portfolio of applications that I mentioned or that are part of their local, institutional framework, and to interconnect them. That interconnection is often referred to as ‘interoperability.’

**Moderator:**

Can you explain more about interoperability and why it's important?

**Dr. Buetow:**

Interoperability has to do with the capacity to share information in a variety of different contexts. In the context of caBIG® it means the ability to share and use information between investigators, between applications, and across the broader infrastructure, the electronic infrastructure that's part of something we call the Cancer Knowledge Cloud.



The importance of interoperability is that one doesn't have to hand-copy or translate information by human intervention. It means that data is captured in a structured way using international standards that permits us to transfer the data, between institutions and organizations and between computer applications using our underlying framework. In that way, the data can be interconnected and the whole can be rendered to mean more than the sum of the individual parts. In other words, it allows the seamless transmission of information between individuals, between organizations, and between institutions. Interoperability also allows the aggregation and use of that information in novel and important ways.

**Moderator:**

Can you give us examples of what caBIG<sup>®</sup> allows researchers to do that they haven't been able to do in the past?

**Dr. Buetow:**

Absolutely. By linking and connecting multiple kinds of information from multiple sources, the members of the cancer community—or the cancer “ecosystem” of everyone who is fighting cancer—can share information and capabilities. They can exchange knowledge that they've acquired through their individual laboratory research or their clinical research studies, as well as share the information out towards the larger community.

caBIG<sup>®</sup>, for example, is being used to enable a scientific endeavor called The Cancer Genome Atlas, or TCGA, being conducted by the National Cancer Institute and the National Human Genome Research Institute. It's a very large-scale project, designed to accelerate our understanding of cancer, and is literally generating terabytes worth of data describing cancer at a comprehensive molecular level. Within this project, there are hundreds of researchers, bioethicists, doctors, nurses, cancer patients, and advocates, who are networked across the country, working collaboratively. To make this work seamlessly requires the ability to collect, analyze and disseminate more than a dozen different types of genomic, laboratory and clinical data coming from dozens of different locations. All that is currently powered by caBIG<sup>®</sup> capabilities, including the network infrastructure called caGrid.

**Moderator:**

Can you give other examples—either for individual researchers or entire institutions—where caBIG<sup>®</sup> plays a role in doing things differently?

**Dr. Buetow:**

Well, the availability of a system like caBIG<sup>®</sup> enables us to contemplate a whole new model of biomedical research. At the most fundamental level, it empowers a researcher to be able to perform operations or to be able to perform analysis or to access information that just wouldn't normally be available to them. We have countless examples of how that's actually working.

Through caBIG<sup>®</sup>, an investigator can just sit down at their desktop, open a web browser, and have access to the most comprehensive molecular information available on the underpinnings of cancer either from the Cancer Genome Atlas project, the project I just described, or from other molecular databases that are plugged in to this caBIG<sup>®</sup>/caGrid network.

caBIG<sup>®</sup> also supports the conduct of clinical trials, both within an individual institution, or to support multi-institutional clinical trials that allow us to develop new therapeutic approaches. For example, in breast cancer, there are studies underway at Duke University to test new treatments, in which caBIG<sup>®</sup> software is being used to help collect patient information and track the impact of those developing treatments on patients with different genetic profiles.



caBIG<sup>®</sup> is also being used to support local, regional, and national efforts to connect and interconnect communities. For instance, the state of Arizona has recently decided to use tools from the caBIG<sup>®</sup> infrastructure—specifically, applications that support biospecimen repository management—to interconnect their institutions across the entire state. In that situation, caBIG<sup>®</sup> is helping to resolve one of the biggest obstacles to productive biomedical research—namely, the difficulty that researchers have in gaining access to large numbers of well characterized and clinically annotated biospecimens.

**Moderator:**

We hear a lot about hospitals and physicians moving towards electronic health records, as well as the idea that individual consumers will soon be using personal health records. What role can caBIG<sup>®</sup> play in this area?

**Dr. Buetow:**

A great deal has been written about the promise of 21st Century medicine. Sometimes it's called "personalized medicine" or "molecular medicine" or "precision medicine." Basically, all these things are about using a great deal of data—biologic and clinical data—in an organized way to determine the best prevention and treatment strategies for an individual.

To achieve this, there's a tremendous amount of information about a person's health that needs to be captured in an organized way, using standard terminologies and vocabularies. Such information can be captured electronically in an electronic medical record used by a physician, or in a personal health record that's accessible to the individual consumer.

Our goal through caBIG<sup>®</sup> is to try to extend the capabilities in this area to the field of cancer. We want to facilitate having an entirely interconnected biomedical universe where all of the patient care encounters are available as a foundation for better health care. That means making knowledge available for biomedical research, making it available to help physicians so that they can see how other patients treated in a similar manner have responded to related treatments, and making it available so that we can identify what's working and not working across the entire health enterprise.

And of course lastly, it makes it available to consumers, so that consumers themselves have the ability to chart their own health directions and have an active role in deciding how to engage in their disease management and treatment strategies.

**Moderator:**

Dr. Buetow, that was very helpful. Thank you very much for your time.

**Dr. Buetow:**

Thank you very much.

**Moderator:**

We'll be back with a special feature on caBIG<sup>®</sup> and the emerging Cancer Knowledge Cloud.